



US012056896B2

(12) **United States Patent**  
**Fragoso Rojas et al.**

(10) **Patent No.:** **US 12,056,896 B2**  
(45) **Date of Patent:** **\*Aug. 6, 2024**

(54) **PRIOR INFORMED POSE AND SCALE ESTIMATION**

(2017.01); *G06T 17/00* (2013.01); *H04N 23/90* (2023.01); *G06T 2207/30244* (2013.01)

(71) Applicant: **Microsoft Technology Licensing, LLC**, Redmond, WA (US)

(58) **Field of Classification Search**  
CPC .. *G06T 7/70*; *G06T 7/80*; *G06T 17/00*; *G06T 2207/30244*

(72) Inventors: **Victor M. Fragoso Rojas**, Bellevue, WA (US); **Mei Chen**, Bellevue, WA (US); **Gabriel Takacs**, Issaquah, WA (US)

See application file for complete search history.

(73) Assignee: **Microsoft Technology Licensing, LLC**, Redmond, WA (US)

(56) **References Cited**

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 37 days.

U.S. PATENT DOCUMENTS

This patent is subject to a terminal disclaimer.

9,031,809 B1 *	5/2015	Kumar .....	<i>G01S 5/163</i>
			702/150
2012/0120199 A1 *	5/2012	Ben Himane .....	<i>G06T 7/75</i>
			348/46
2015/0172626 A1 *	6/2015	Martini .....	<i>G06T 7/246</i>
			348/50

OTHER PUBLICATIONS

(21) Appl. No.: **17/931,238**

Sweeney, C., Fragoso, V., Höllerer, T., & Turk, M. (2014). gdl: A scalable solution to the generalized pose and scale problem. In Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, Sep. 6-12, 2014, Proceedings, Part IV 13 (pp. 16-31). Springer International Publishing.\*

(22) Filed: **Sep. 12, 2022**

\* cited by examiner

(65) **Prior Publication Data**

US 2024/0087155 A1 Mar. 14, 2024

**Related U.S. Application Data**

*Primary Examiner* — Sarah Lhymn

(63) Continuation of application No. 16/744,068, filed on Jan. 15, 2020, now Pat. No. 11,443,455.

(74) *Attorney, Agent, or Firm* — Alleman Hall & Tuttle LLP

(60) Provisional application No. 62/925,605, filed on Oct. 24, 2019.

(57) **ABSTRACT**

(51) **Int. Cl.**

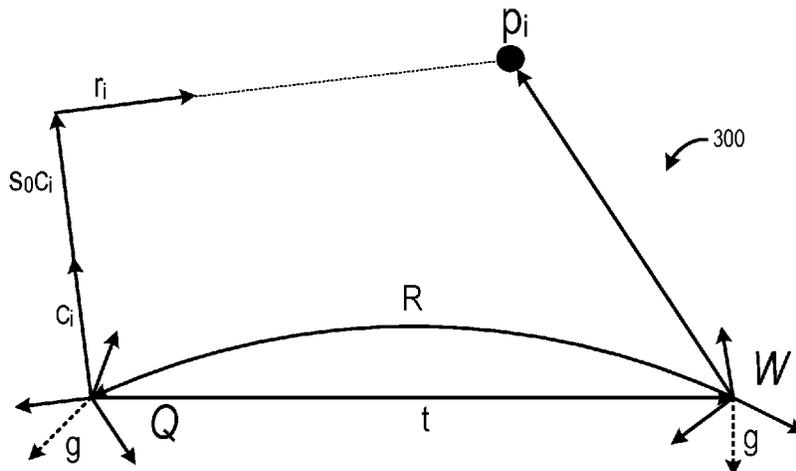
<i>G06T 7/73</i>	(2017.01)
<i>G06T 7/60</i>	(2017.01)
<i>G06T 7/70</i>	(2017.01)
<i>G06T 7/80</i>	(2017.01)
<i>G06T 17/00</i>	(2006.01)
<i>H04N 23/90</i>	(2023.01)

A scale and pose estimation method for a camera system is disclosed. Camera data for a scene acquired by the camera system is received. A scale prior parameter characterizing scale of the camera system is received. A cost of a cost function is calculated for a similarity transformation. The cost of the cost function is influenced at least by the scale prior parameter. Based at least on the cost function being less than a threshold cost, an estimated scale and pose of the camera system is output based on the similarity transformation.

(52) **U.S. Cl.**

CPC ..... *G06T 7/75* (2017.01); *G06T 7/60* (2013.01); *G06T 7/70* (2017.01); *G06T 7/80*

**20 Claims, 5 Drawing Sheets**



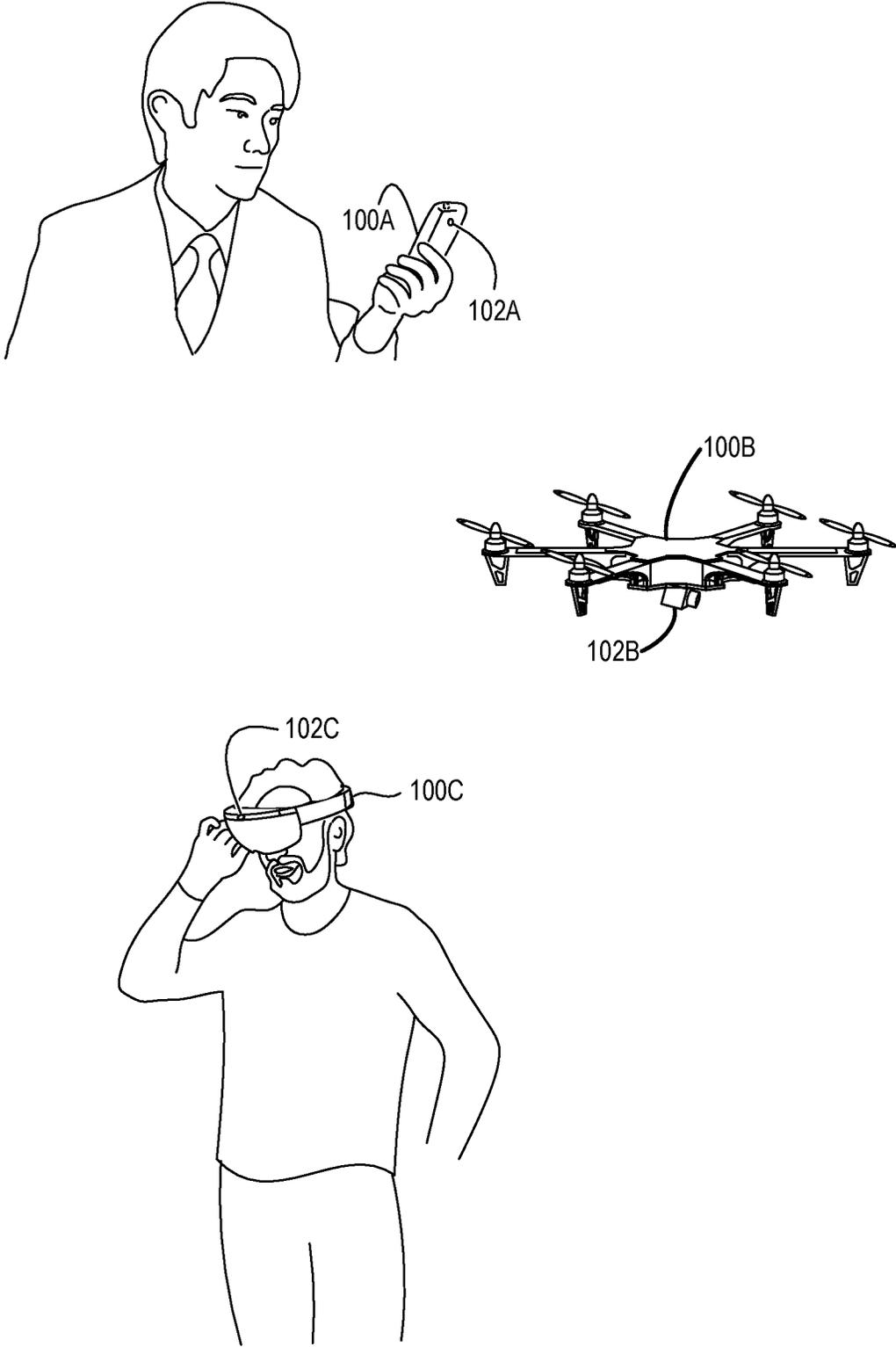


FIG. 1

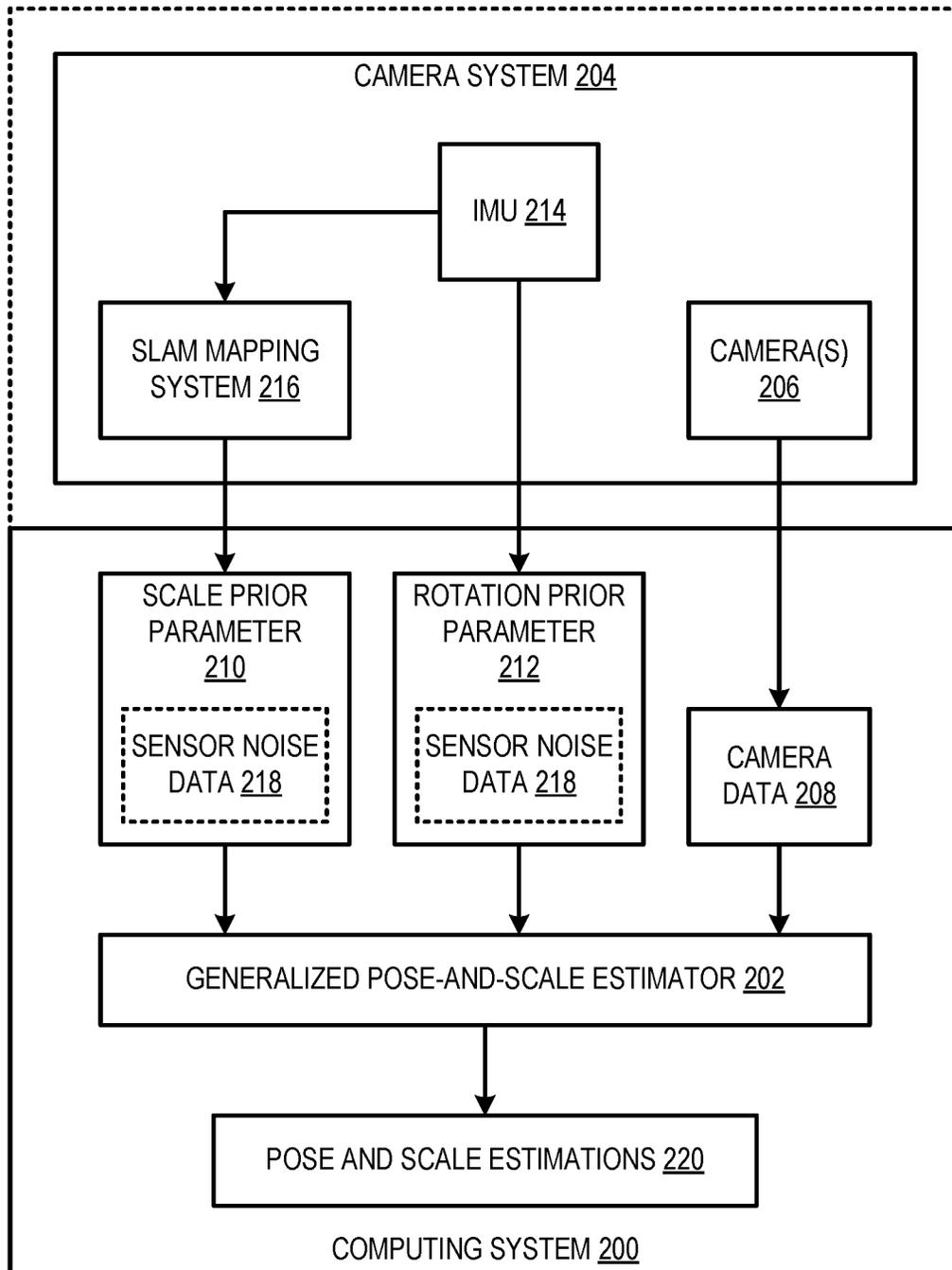


FIG. 2

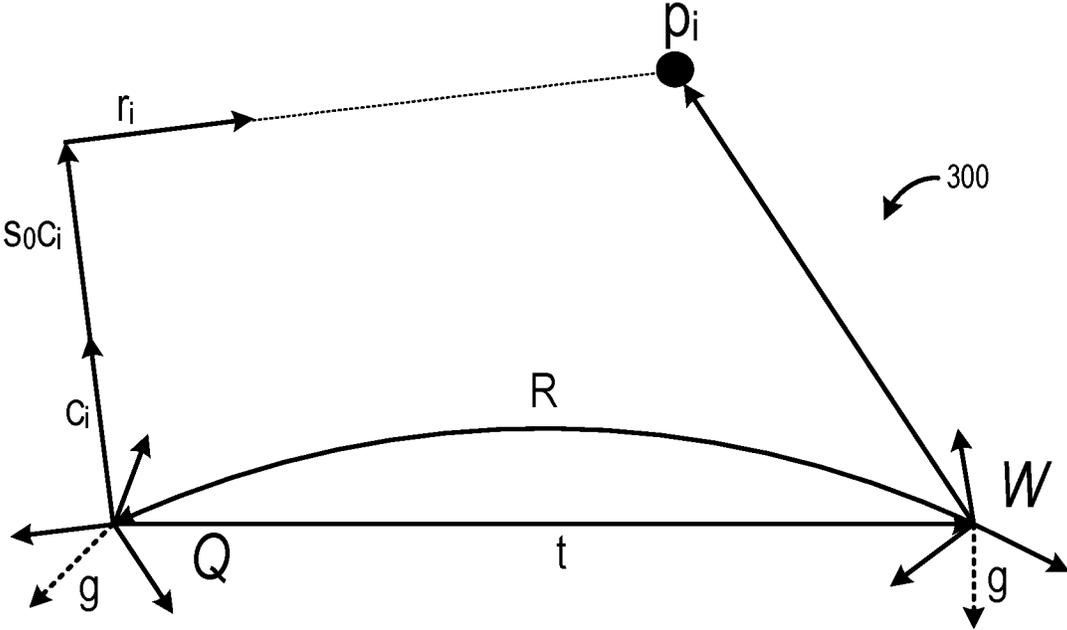


FIG. 3

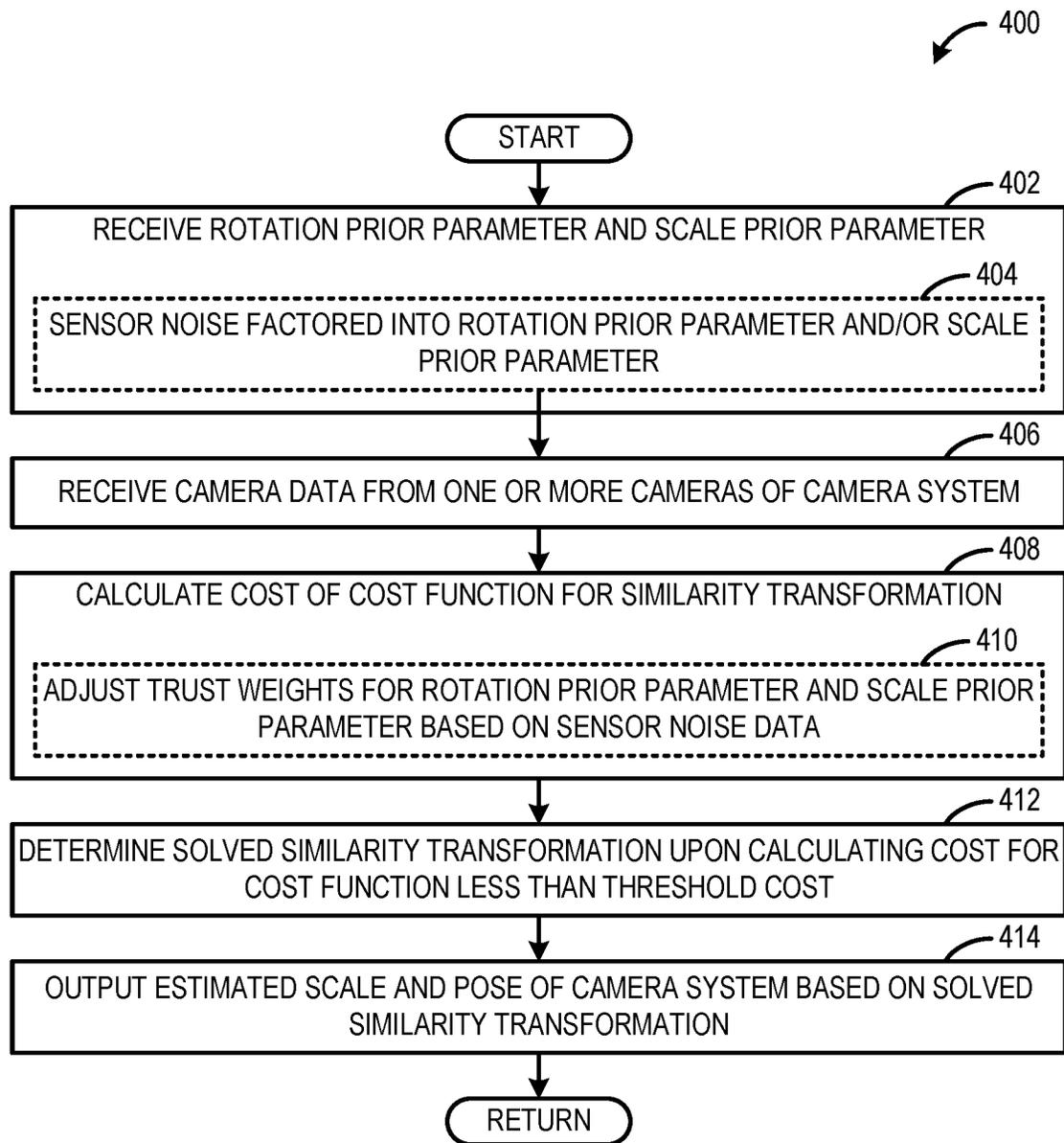


FIG. 4

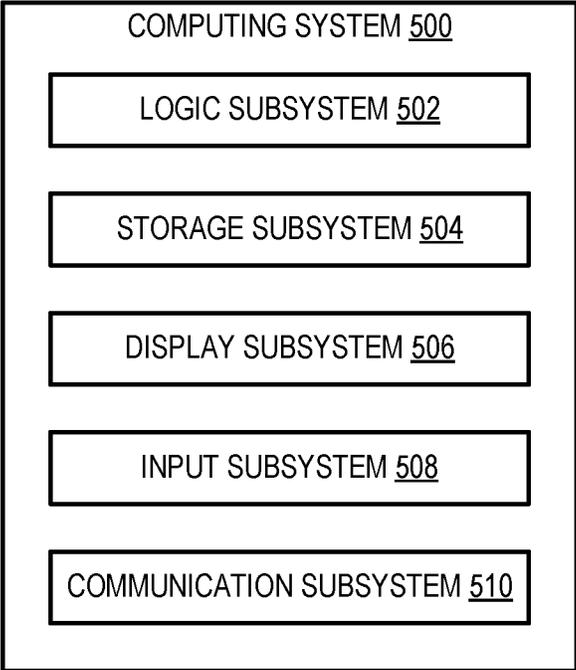


FIG. 5

## PRIOR INFORMED POSE AND SCALE ESTIMATION

### CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. Non-Provisional patent application Ser. No. 16/744,068, filed Jan. 15, 2020, which claims priority to U.S. Provisional Patent Application No. 62/925,605, filed Oct. 24, 2019, the entirety of each of which is hereby incorporated herein by reference.

### BACKGROUND

Camera pose estimation, i.e., estimating the position and orientation of a camera system relative to a scene, is a central step in computer vision. For example, pose estimation may be utilized in computer vision applications, such as Simultaneous Localization and Mapping (SLAM), visual localization, augmented reality (AR), 3D mapping, and robotics. Such camera pose estimation may enable spatial registration between a camera coordinate system and a world coordinate system of the scene.

### SUMMARY

A scale and pose estimation method for a camera system is disclosed. Camera data for a scene acquired by one or more cameras of the camera system is received. A rotation prior parameter characterizing a gravity direction is received. A scale prior parameter characterizing scale of the camera system is received. A cost of a cost function is calculated for a similarity transformation that is configured to encode a scale and pose of the camera system. The cost of the cost function is selectively influenced by the rotation prior parameter and the scale prior parameter. A solved similarity transformation is determined upon calculating a cost for the cost function that is less than a threshold cost. An estimated scale and pose of the camera system is output based on the solved similarity transformation.

This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter. Furthermore, the claimed subject matter is not limited to implementations that solve any or all disadvantages noted in any part of this disclosure.

### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows different example electronic devices including camera systems used to spatially register the electronic devices with real-world environments.

FIG. 2 shows an example pose and scale estimator implemented by a computing system.

FIG. 3 diagrammatically shows an example transformation of a world coordinate system to a camera coordinate system.

FIG. 4 is a flowchart of an example method for estimating pose and scale for a camera system.

FIG. 5 schematically shows an example computing system.

### DETAILED DESCRIPTION

Many real-world applications in computer vision such as augmented reality (AR), three-dimensional (3D) mapping,

and robotics require both fast and accurate estimation of camera pose and scale. Achieving high speed and maintaining high accuracy when estimating pose and scale of a camera system are often conflicting goals. Some conventional approaches for estimating scale and pose of a camera system produce estimations at a high speed, however such estimations may have lower accuracy. Other conventional approaches for estimating scale and pose of a camera system produce estimations with high accuracy, however such estimations may be produced at lower speeds that may not be feasible for many applications.

The present description is directed to an approach for estimating scale and absolute pose of a camera system in a manner that achieves both high speed and high accuracy. In particular, this approach exploits a priori knowledge of the solution by using rotation and scale prior parameters to bias the solution to more quickly obtain accurate estimations. Use of such rotation and scale prior parameters may accelerate the estimation process and improve scale and pose estimation accuracy relative to other approaches that do not use scale and rotation prior parameters.

Furthermore, in some implementations, the estimation approaches described herein may be configured to operate with selective and/or varying influence of the scale and rotation prior parameters on scale/pose estimation. For example, the contribution of each prior parameter may be flexibly weighed based on noisy input data, for example as might occur due to sensor noise. As such, the estimation approach can be robust to noisy sensor inputs, which can arise in AR, robotics, and other applications.

FIG. 1 shows aspects of different examples of electronic devices (100A-C) that each have a camera system (102A-C) that may employ scale and pose estimation. Device 100A is a smartphone that includes a camera system 102A. For example, the camera system 102A may be used to present an augmented reality experience on a display of the smartphone. Device 100B is an automated flying vehicle that includes a camera system 102B. For example, the camera system 102B may be used to provide automated flight control of the automated flying vehicle. Device 100C is a virtual-reality or augmented-reality headset that includes a camera system 102C. For example, the camera system 102A may be used to present a virtual reality or augmented reality experience on a display of the headset. The example scale and pose estimation approaches disclosed herein may be applicable to these and other camera systems.

FIG. 2 schematically shows an example computing system 200 configured to implement a generalized pose-and-scale estimator (or solver) 202. The estimator 202 may be implemented by the computing system as software, hardware, firmware, or any combination thereof. The estimator 202 may be configured to output pose and scale estimations for a camera system 204 including one or more cameras 206 based on camera data 208 received from the camera system 204. For example, the camera data 208 may include camera measurements of a collection of 3D points of a 3D model of a scene that are captured by one or more cameras of the camera system 204. In some examples, the estimator 202 may receive camera data in the form of an image or images from the camera system 204, and process the images to form a collection of 3D points of a 3D model of a scene. The estimator 202 may be configured to perform any suitable image processing operations to obtain camera data suitable for estimating scale and pose estimations of the camera system 204.

In some implementations, the camera system 204 may include a plurality of different cameras having different

positions fixed relative to one another, and the camera data 208 may include data from each of the different cameras. Alternatively, in some implementations, the camera data 208 may include data from a single moving camera of the camera system 204 over time. The camera data 208 may include any suitable form of data derived from images captured by the camera(s) 206 of the camera system 204.

The estimator 202 further may be configured to receive a scale prior parameter 210 and a rotation prior parameter 212 as input. The scale prior parameter 210 characterizes scale of the camera system. The rotation prior parameter 212 characterizes a gravity direction. The scale prior parameter 210 and the rotation prior parameter 212 may inform the estimator 202 of behavior of the camera system 204 that may influence current estimations of scale and pose output by the estimator 202. These prior parameters may be derived from sensor data of the camera system 204. For example, the rotation prior parameter 212 can be derived from a gravity direction using measurements from one or more inertial sensors of an inertial measurement unit (IMU) 214. The IMU 214 may include one or more accelerometers, gyroscopes, and/or other motion sensors that are configured to provide an indication of gravity direction. Additionally, as an example, the scale prior parameter 210 can be obtained from a simultaneous localization and mapping (SLAM) system 216 of the camera system 204. The SLAM system 216 may be configured to keep a unit scale utilizing data from the IMU 214. The scale prior parameter 210 and the rotation prior parameter 212 may be provided from any suitable sensor of the camera system 204. In some examples, the scale prior parameter 210 and the rotation prior parameter 212 may be received by the estimator 202 prior to corresponding camera data 208, such that the scale prior parameter 210 and the rotation prior parameter 212 provide hints or suggestions of pose and orientation information for the camera system 204. In other examples, the scale prior parameter 210 and the rotation prior parameter 212 may be received by the estimator 202 substantially at the same time as the corresponding camera data 208. In yet other examples, the scale prior parameter 210 and the rotation prior parameter 212 may be received by the estimator 202 after the corresponding camera data 208 has been received by the estimator 202.

In some implementations, sensor noise data 218 optionally may be factored into the scale prior parameter 210 and/or the rotation prior parameter 212. The sensor noise data 218 may provide a quantifiable indication of sensor noise in operation of the IMU 214. In some examples, sensor noise data may correspond to operation of the IMU when the scale prior parameter 210 and the rotation prior parameter 212 are determined. In other words, the sensor noise data 218 may provide an indication of the reliability or accuracy of the scale prior parameter 210 and the rotation prior parameter 212 to the estimator 202.

The estimator 202 is configured to use the input data (e.g., camera data, scale and rotation prior parameters) to solve an optimization problem to estimate a probable pose/scale of the camera system 204. In one example, the estimator 202 is configured to calculate a cost of a cost function for a similarity transformation that is configured to encode a scale and pose of the camera system 204.

FIG. 3 diagrammatically shows an example similarity transformation 300 for which the estimator 202 shown in FIG. 2 may be configured to solve. The similarity transformation may align a world coordinate system (W) and a camera coordinate system (Q) by estimating a rotation (R), translation (t), and scale(s) of the coordinate systems relative

to one another. The inputs to the estimator include 1) camera data, for example a set of three dimensional points ( $p_i$ ) in the world coordinate system, 2) a set of camera positions (positioned at the end of the ray  $s_0c_i$ ) in the multi-camera coordinate system, and 3) the gravity vectors ( $g$ ) in the world coordinate system and the multi-camera coordinate system. In this example, the scale prior ( $s_0$ ) influences the length of the ray to properly locate the camera positions (ray length for proper positioning and adjusted scale= $s_0c_i$ ). The non-adjusted camera position(s) are denoted by ray ( $c_i$ ). The ray ( $r_i$ ) extends from the point of the camera position in the direction of the point ( $p_i$ ).

Returning to FIG. 2, the estimator 202 may be configured to solve the similarity transformation by calculating a cost of a cost function for the similarity transformation. For example, the cost function may include a generalized least squares cost function. The cost function used by the estimator 202 may be a modification of a generalized least squares cost function shown as Equation (1) shown below. Equation (1) is a generalized least-squares cost function that may be used to estimate pose and scale of a non-central camera. Equation (1) may be used to solve for the rotation (R), the translation (t), and the scale (s) that minimizes the error/cost. Equation (1) does not include rotation and scale priors

$$J(R,t,s,\alpha)=\sum_{i=1}^n\|\alpha_i r_i-(Rp_i+t-sc_i)\|^2 \quad (1)$$

where  $r_i$  is a unit-vector indicating the direction from the position of the camera  $c_i$  to a 3D point  $p_i$ ;  $\alpha_i$  is the depth of the point  $p_i$  with respect to the camera position  $c_i$ ;  $\alpha$  is a vector holding the depths;  $R \in SO(3)$  is a rotation matrix;  $t \in \mathbb{R}^3$  is a translation vector; and  $s \in \mathbb{R}$  is the scale. The pose-and-scale formulation shown in Eq. (1) accumulates the errors between the transformed  $i$ -th 3D point ( $Rp_i+t-sc_i$ ) and the same point described with respect to the camera  $\alpha_i r_i$ . The rotation R, the translation t, and the scaled camera position  $sc_i$  transform a 3D point from a world coordinate system to the coordinate system of the generalized camera.

In order to find the minimizer ( $R^*$ ;  $t^*$ ;  $s^*$ ;  $\alpha^*$ ), Equation 1 is rewritten as a function that only depends on the rotation matrix. In this step, the translation t, scale s, and depth  $\alpha_i$  can be written as a linear function of the rotation matrix R. Thus, it is possible to re-write the pose-and-scale least-squares cost formulation of Equation (1) as Equation (2) shown below:

$$J(R)=\sum_{i=1}^n\|\alpha_i(R)r_i-(Rp_i-t(R)-s(R)c_i)\|^2=\text{vec}(R)^T M \text{vec}(R) \quad (2)$$

where  $\text{vec}(R)$  is a vectorized form of the rotation matrix, and M is a squared matrix capturing the constraints from the input 2D-3D correspondences; the dimensions of M depend on the vectorization and representation of  $\text{vec}(R)$ . Given the cost function  $J(R)$ , the optimal rotation  $R^*$  is found by solving a polynomial system representing the constraint that the gradient  $\nabla_q J(R^*)=0$  is null with respect to the rotation-quaternion parameters q, and rotation-parameter constraints (e.g., ensuring a unit-norm quaternion).

Note that although the cost function is discussed herein in terms of solving for the minimizer, the cost function in some examples may be solved such that the cost is below a designated threshold cost.

In order to impose scale and rotation priors that influence the cost equation, Equation (1), is modified to include regularizer terms. Adding these regularizer terms leads to the least-squares cost function of Equation (3) shown below:

$$J'=J(R,t,s,\alpha)+\lambda_s(s_0-s)^2+\lambda_g\|g_Q \times Rg_W\|^2 \quad (3)$$

where  $s_0$  is the scale prior;  $g_Q$  and  $g_W$  are the gravity directions of the multi-camera setting and world, respec-

## 5

tively; the symbol  $\times$  represents the cross-product operator; and  $\lambda_s$  and  $\lambda_g$  are weights controlling the contribution of the scale and rotation priors, respectively. These weights (i.e.,  $\lambda_s$  and  $\lambda_g$ ) may be greater than or equal to zero.

These priors bias the cost function to penalize scale and rotation estimates that deviate from prior estimates. In particular, the scale prior regularizer term is an additional parameter that penalizes deviation from a prior assumption about scale. The scale is the reduction or enlargement factor that the points described in the camera coordinate system need to match the metric units of those points described in the world coordinate system. The scale prior regularizer term  $\lambda_s(s_0-s)^2$  is configured to influence the cost by imposing a penalty on the cost when the queried scale  $s$  deviates from the scale prior  $s_0$ . On the other hand, the rotation prior is an additional parameter that captures a notion of difference between a candidate solution and a prior assumption about gravity direction. The rotation prior regularizer term  $\lambda_g\|g_Q \times Rg_w\|^2$  is configured to influence the cost by imposing a misalignment penalty on the cost when the transformed world gravity direction  $Rg_w$  and the query gravity direction  $g_Q$  are misaligned.

In some implementations, the estimator **202** may be configured to control the influence of the scale prior parameter and the rotation prior parameter on the cost function that determines the pose and scale estimations. In one example, each prior may have a corresponding scalar or trust weight (i.e.,  $\lambda_s$  and  $\lambda_g$ ) that may be set based on sensor noise data **218** of the IMU sensor **214** from which the prior was obtained. In some implementations, the weights may be predetermined based on empirical sensor data. In other implementations, the weights may be dynamically adjusted based on the sensor noise data **218**. For example, as sensor noise dynamically increases, the weight dynamically decreases such that the corresponding prior parameter has less (or no) influence on the cost of the cost function and further the estimations **220**. Likewise, as sensor noise dynamically decreases, the weight dynamically increases such that the corresponding prior parameter has greater influence on the cost of the cost function and further the estimations **220**.

In order to solve for rotation/pose and scale, the cost  $J$  may be re-written as a function that only depends on the rotation matrix. To do so, it is mathematically convenient to use Equation (4) shown below.

$$x = [\alpha_1 \dots \alpha_n s^T]^T. \quad (4)$$

The gradient evaluated at the optimal  $x^*$  must satisfy the following constraint:  $\nabla_x J|_{x=x^*} = 0$ . From this constraint, the relationship shown in Equation (5) shown below is defined as

$$x = (A^T A + P)^{-1} A^T W b + (A^T A + P)^{-1} P x_0 = \begin{bmatrix} U \\ S \\ V \end{bmatrix} W b + \lambda_s s_0 l, \quad (5)$$

where

$$A = \begin{bmatrix} r_1 & & c_1 & -I \\ & \ddots & \vdots & \vdots \\ & & r_n c_n & -I \end{bmatrix}, b = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} \quad (6)$$

$$P = \begin{bmatrix} 0_{n \times n} & & & \\ & \lambda_s & & \\ & & 0_{3 \times 3} & \\ & & & R \end{bmatrix}, W = \begin{bmatrix} R & & \\ & \ddots & \\ & & R \end{bmatrix}$$

and

$$x_0 = [0_n^T s_0 0_3^T]^T.$$

## 6

$(A^T A + P)^{-1} A^T$  is partitioned into three matrices  $U$ ,  $S$ , and  $V$  such that the depth, scale, and translation parameters are functions of  $U$ ,  $S$ , and  $V$ , respectively. These matrices and the vector  $l$  can be computed in closed form by exploiting the sparse structure of the matrices  $A$  and  $P$ .

Equation (5) provides a linear relationship between the depth, scale, and translation and the rotation matrix. Consequently, these parameters are computed as a function of the rotation matrix in Equation (7) shown below

$$\begin{aligned} \alpha_i(R) &= u_i^T W b + \lambda_s s_0 l_i \\ s(R) &= S W b + \lambda_s s_0 l_{n+1} \\ t(R) &= V W b + \lambda_s s_0 l_j \end{aligned} \quad (7)$$

where  $u_i^T$  is the  $i$ -th row of matrix  $U$ ,  $l_j$  is the  $j$ -th entry of the vector  $l$ , and  $l_j$  corresponds to the last three entries of the vector  $l$ .

In order to re-write the regularized least-squares cost function (i.e., Equation (3)) as clearly as possible, Equation (8) is formulated as shown below.

$$\begin{aligned} e_i &= a_i(R) r_i - (R p_i + t(R) - s(R) c_i) = n_i + k_i \\ n_i &= u_i^T W b r_i - R p_i - V W b + S W b c_i \\ k_i &= \lambda_s s_0 (l_i r_i - l_j^T l_{n+1} c_i) \end{aligned} \quad (8)$$

The residual  $e_i$  is divided into two terms:  $\eta_i$ , the residual part considering the unconstrained terms; and  $k_i$ , the residual part considering the scale-prior-related terms. Note again that when  $\lambda_s = 0$ ,  $k_i$  becomes null and  $e_i$  becomes the residual corresponding to the original cost function (i.e., Equation (1)).

Using the definitions from Equation (8) and the scale, depth, and translation relationships shown in Equation (7), the regularized least-squares cost function shown in Equation (3) can be re-written as Equation (9) shown below

$$J = J'_{gDLS} + J'_s + J'_g = \text{vec}(R)^T M \text{vec}(R) + 2d^T \text{vec}(R) + k, \quad (9)$$

where

$$\begin{aligned} J'_{gDLS} &= \sum_{i=1}^n e_i^T e_i = \sum_{i=1}^n n_i^T n_i + 2k_i^T n_i + k_i^T k_i = \\ & \text{vec}(R)^T M_{gDLS} \text{vec}(R) + 2d_{gDLS}^T \text{vec}(R) + k_{gDLS} \\ J'_s &= \lambda_s (s_0 - S(R))^2 = \text{vec}(R)^T M_{sS} \text{vec}(R) + 2d_s^T \text{vec}(R) + k_s \\ J'_g &= \lambda_g \|g_Q \times Rg_w\|^2 = \text{vec}(R)^T M_g \text{vec}(R) \\ M &= M_{gDLS} + M_s + M_g \\ d &= d_{gDLS} + d_s \\ k &= k_{gDLS} + k_s. \end{aligned} \quad (10)$$

The parameters of Equation (9) (i.e.,  $M_{gDLS}$ ,  $M_s$ ,  $M_g$ ,  $d_{gDLS}$ ,  $d_s$ ,  $k_{gDLS}$ , and  $k_s$ ) can be computed in closed form and in  $O(n)$  time. Equation (9) generalizes the unconstrained quadratic function shown in Equation (1) when both scale and rotation priors are disabled, i.e.,  $\lambda_g = \lambda_s = 0$ , then  $J(R) = J(R)$ . Also, note that the weights  $\lambda_g$  and  $\lambda_s$  control the contribution of each of the priors independently. Such independent control allows the estimator **202** to flexibly adapt the cost function to many scenarios. For instance, these weights can be adjusted so that the cost function reflects the confidence of certain priors based on sensor noise, reduces the effect of noise present in the priors, and fully disables one or both of the priors from influencing the cost of the cost function.

7

Equations (4)-(8) may be used as a mechanism to rewrite the Equation (3) to only depend on a rotation matrix that is encoded in Equations (9) and (10). In particular, Equations (5)-(6) are used to compute depth, scale, and translation as a linear transformation of rotation.

Given that the prior-based pose-and-scale cost function (i.e., Equation (3)) depends only on the rotation matrix, R can be solved to minimize Equation (9). To achieve this step, the rotation matrix R can be represented using a quaternion  $q=[q_1 \ q_2 \ q_3 \ q_4]$ . To compute all the minimizers of Equation (9), a polynomial system may be built so that it encodes the first-order optimality conditions and the unit-norm-quaternion constraint shown below in Equation (11):

$$\begin{cases} \frac{\partial J'}{\partial q_j} = 0, & \forall_j = 1, \dots, 4 \\ q_j(q^T q - 1) = 0, & \forall_j = 1, \dots, 4 \end{cases} \quad (11)$$

The polynomial system on q shown in Equation (11) encodes the unit-norm-quaternion constraint within Equation (12) shown below.

$$\frac{\partial(q^T q - 1)^2}{\partial q_j} = q_j(q^T q - 1) = 0, \forall_j. \quad (12)$$

Equation (12) yields efficient elimination templates and small action matrices, which delivers efficient polynomial solvers. The efficient polynomial solver adopted by the cost function leverages the rotation representation shown in Equation (13) shown below.

$$\text{vec}(R)=[q_1^2 q_2^2 q_3^2 q_4^2 q_1 q_2 q_1 q_3 q_1 q_4 q_2 q_3 q_2 q_4 q_3 q_4]^T. \quad (13)$$

Given this representation, the dimensions of the parameters of the regularized least-squares cost function shown in Equation (9) become  $M \in \mathbb{R}10 \times 10$ ,  $d \in \mathbb{R}10$ , and  $k \in \mathbb{R}$ .

The cost function efficiently yields eight rotations after solving the polynomial system stated in Equation (11). After computing these rotations, the estimator **202** discards quaternions with complex numbers whose imaginary part falls below a certain threshold, and then recovers the depth, scale, and translation using Equation (7). Finally, the estimator **202** uses the computed similarity transformations to discard solutions that map the input 3D points behind any camera. By solving for the rotation matrix using Equations (9) and (10), the pose and scale may be estimated in a more quick and accurate manner than using Equation (2), which does not include rotation and scale priors.

As discussed above, the scale prior parameter **210** and the rotation prior parameter **212** may be configured to influence the cost of the cost function. The estimator **202** may be configured to perform the calculation to optimize at least a rotation term and a translation term of the similarity transformation. The estimator **202** may be further configured to perform the calculation to optimize the scale and depth terms as well, in some examples.

In some examples, the estimator **202** may determine a solved similarity transformation upon calculating a cost for the cost function that is less than a threshold cost. The threshold cost may be set to any suitable cost to provide a suitably accurate estimation of scale and pose in a suitably quick timeframe. In some examples, the threshold cost may correspond to a minimized cost calculated for the cost function as discussed above.

8

Further, the estimator **202** may be configured to output the estimated scale and pose **220** of the camera system **204** based on the solved similarity transformation. The similarity transformation may include one or more rigid transformations (e.g., rotation, translation) followed by a dilation (e.g., reduce/enlarge by a scale factor). Such a similarity transformation (and/or corresponding output of the estimator **202**) may be used to align two or more reconstructions of a 3D model of a scene (e.g., 3D point cloud) computed with different sensors, or to localize a rig of cameras within an existing 3D reconstruction, among other scenarios. The estimator **202** may be configured to output the pose and scale estimations **220** to any suitable processing logic of the camera system **204**, the computing system **200**, and/or any other suitable device.

In some implementations, the computing system **200** may include both the estimator **202** and the camera system **204** incorporated into a same device. For example, the estimator and the camera system may be incorporated into an augmented reality device worn by a user. In another example, the estimator and the camera system may be incorporated into a robotic device configured for autonomous movement. In other implementations, the camera system **204** may be incorporated into a different device than the estimator **202**. For example, the computing system **200** may be remote from the camera system **204** and connected via a computer network. In some examples, the estimator **202** may be implemented by a service computing system that is configured to receive the scale and rotation priors and camera data from the camera system, and return the pose and scale estimations to the camera system or an associated computing device. In some examples, the service may be a hologram sharing service for augmented reality devices. The hologram sharing service may maintain a reference map, and a client camera system may send input data to the hologram sharing service such that the hologram sharing service can use the input data to localize queried images with the reference map. The hologram sharing service may send the calculated pose and scale estimations back to the client camera system so that holograms can be positioned correctly on an augmented reality display associated with the client camera system in a quick and accurate manner.

FIG. 4 shows an example method **400** for estimating pose and scale of a camera system including one or more cameras. For example, the method **400** may be performed by the estimator **202** shown in FIG. 2. At **402**, rotation and scale prior parameters are received. The rotation prior parameter characterizes a gravity direction. The scale prior parameter may characterize scale, depth, and/or translation of the camera system. In some examples, the rotation prior parameter and the scale prior parameter are derived from measurements of one or more inertial sensors of the camera system or an associated computing device.

In some implementations where the estimator is configured to dynamically adjust the contribution of the rotation and scale prior parameters based on sensor noise of one or more inertial sensors, at **404**, sensor noise data optionally may be factored into the scale prior parameter and/or the rotation prior parameter. At **406**, camera data for a scene is received from the one or more cameras of the camera system. In some examples, the camera data may include a collection of three-dimensional (3D) points of a 3D model of the scene. In some examples, the camera data may include whole or partial image frames of the scene that are optionally processed to form a 3D point cloud of the scene.

In some examples, the camera system may include a plurality of different cameras having different positions fixed

relative to each other and the camera data may correspond to each of the different cameras. In other examples, the camera system may include a single camera movable throughout the scene and the camera data may correspond to the single camera at different positions within the scene.

At **408**, a cost of a cost function for a similarity transformation that is configured to encode a scale and pose of the camera system is calculated. In one example, the cost may be calculated using Equations (9) and (10) shown described above. Such calculations may be performed to optimize a rotation term and a translation term of the similarity transformation. Further, the cost function includes the rotation prior parameter and the scale prior parameter such that the cost of the cost function is selectively influenced by the rotation prior parameter and the scale prior parameter. In some examples, the rotation prior parameter and the scale prior parameter selectively influence the cost of the cost function by selectively imposing a penalty on the cost of the cost function based on queried input data differing from prior data.

In some implementations, the cost function may include a rotation weight corresponding to the rotation prior parameter and a scale weight corresponding to the scale prior parameter. In such implementations, at **410**, optionally the rotation weight and the scale weight may be dynamically adjusted based on sensor noise data factored into the scale prior parameter and/or the rotation prior parameter. As an example, the rotation weight and the scale weight may be dynamically decreased as sensor noise increases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have less influence on the cost function as sensor noise increases. As another example, the rotation weight and the scale weight may be dynamically increased as sensor noise decreases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have more influence on the cost function as sensor noise decreases. In some examples, either or both of the rotation weight and the scale weight may be dynamically set to zero based on sensor noise being greater than a threshold noise level such that the corresponding rotation prior parameter and the corresponding scale prior parameter have no influence on the cost function.

At **412**, a solved similarity transformation is determined upon calculating a cost for the cost function that is less than a threshold cost. The threshold cost may be set to any suitable cost. In some examples, the threshold cost is a minimized cost for the cost function. At **414**, an estimated scale and pose of the camera system is output based on the solved similarity transformation. The method **400** may be performed for each camera (i) of a plurality of cameras of a multi-camera system (or a single moving camera).

The herein described pose and scale estimation method can be broadly applicable to a variety of multi-camera or moving camera applications, such as in augmented reality and robotics applications. For instance, when trying to align a 3D reconstruction A (e.g., computed via a structure-from-motion algorithm) and a reconstruction B (e.g., computed with a SLAM system), the pose and scale method approach can be used to estimate a transformation that can align both reconstructions exploiting sensor information. This alignment reveals pose and scale information that enables accurate AR renderings and/or to localize precisely the position and orientation of a mobile device within an environment. The rotation and scale prior parameters provided by external sensors can be selectively used an input to quickly and accurately estimate pose and scale of a camera system. In some implementations, the contributions of these rotation

and scale prior parameters can be selectively reduced or minimized based on the sensor noise of the sensor from which the prior parameters were obtained. In this way, the estimation method may be robust to noisy scale and gravity input data.

The methods and processes described herein may be tied to a computing system of one or more computing devices. In particular, such methods and processes may be implemented as an executable computer-application program, a network-accessible computing service, an application-programming interface (API), a library, or a combination of the above and/or other compute resources.

FIG. 5 schematically shows a simplified representation of a computing system **500** configured to provide any to all of the compute functionality described herein. Computing system **500** may take the form of one or more personal computers, network-accessible server computers, tablet computers, home-entertainment computers, gaming devices, mobile computing devices, mobile communication devices (e.g., smart phone), virtual/augmented/mixed reality computing devices, wearable computing devices, Internet of Things (IoT) devices, embedded computing devices, camera system, robotic device, autonomous vehicle, and/or other computing devices. For example, computing system **500** may take the form of computing system **200** and/or camera system **204** shown in FIG. 2.

Computing system **500** includes a logic subsystem **502** and a storage subsystem **504**. Computing system **500** may optionally include a display subsystem **506**, input subsystem **508**, communication subsystem **510**, and/or other subsystems not shown in FIG. 5.

Logic subsystem **502** includes one or more physical devices configured to execute instructions. For example, the logic subsystem may be configured to execute instructions that are part of one or more applications, services, or other logical constructs. The logic subsystem may include one or more hardware processors configured to execute software instructions. Additionally or alternatively, the logic subsystem may include one or more hardware or firmware devices configured to execute hardware or firmware instructions. Processors of the logic subsystem may be single-core or multi-core, and the instructions executed thereon may be configured for sequential, parallel, and/or distributed processing. Individual components of the logic subsystem optionally may be distributed among two or more separate devices, which may be remotely located and/or configured for coordinated processing. Aspects of the logic subsystem may be virtualized and executed by remotely-accessible, networked computing devices configured in a cloud-computing configuration.

Storage subsystem **504** includes one or more physical devices configured to temporarily and/or permanently hold computer information such as data and instructions executable by the logic subsystem. When the storage subsystem includes two or more devices, the devices may be collocated and/or remotely located. Storage subsystem **504** may include volatile, nonvolatile, dynamic, static, read/write, read-only, random-access, sequential-access, location-addressable, file-addressable, and/or content-addressable devices. Storage subsystem **504** may include removable and/or built-in devices. When the logic subsystem executes instructions, the state of storage subsystem **504** may be transformed—e.g., to hold different data.

Aspects of logic subsystem **502** and storage subsystem **504** may be integrated together into one or more hardware-logic components. Such hardware-logic components may include program- and application-specific integrated circuits

(PASIC/ASICs), program- and application-specific standard products (PSSP/ASSPs), system-on-a-chip (SOC), and complex programmable logic devices (CPLDs), for example.

The logic subsystem and the storage subsystem may cooperate to instantiate one or more logic machines. As used herein, the term “machine” is used to collectively refer to the combination of hardware, firmware, software, instructions, and/or any other components cooperating to provide computer functionality. In other words, “machines” are never abstract ideas and always have a tangible form. A machine may be instantiated by a single computing device, or a machine may include two or more sub-components instantiated by two or more different computing devices. In some implementations a machine includes a local component (e.g., software application executed by a computer processor) cooperating with a remote component (e.g., cloud computing service provided by a network of server computers). The software and/or other instructions that give a particular machine its functionality may optionally be saved as one or more unexecuted modules on one or more suitable storage devices.

Non-limiting examples of training procedures for adjusting trainable parameters include supervised training (e.g., using gradient descent or any other suitable optimization method), zero-shot, few-shot, unsupervised learning methods (e.g., classification based on classes derived from unsupervised clustering methods), reinforcement learning (e.g., deep Q learning based on feedback) and/or generative adversarial neural network training methods, belief propagation, RANSAC (random sample consensus), contextual bandit methods, maximum likelihood methods, and/or expectation maximization. In some examples, a plurality of methods, processes, and/or components of systems described herein may be trained simultaneously with regard to an objective function measuring performance of collective functioning of the plurality of components (e.g., with regard to reinforcement feedback and/or with regard to labelled training data). Simultaneously training the plurality of methods, processes, and/or components may improve such collective functioning. In some examples, one or more methods, processes, and/or components may be trained independently of other components (e.g., offline training on historical data).

When included, display subsystem **506** may be used to present a visual representation of data held by storage subsystem **504**. This visual representation may take the form of a graphical user interface (GUI). Display subsystem **506** may include one or more display devices utilizing virtually any type of technology. In some implementations, display subsystem may include one or more virtual-, augmented-, or mixed reality displays.

When included, input subsystem **508** may comprise or interface with one or more input devices. An input device may include a sensor device or a user input device. Examples of user input devices include a keyboard, mouse, touch screen, or game controller. In some embodiments, the input subsystem may comprise or interface with selected natural user input (NUI) componentry. Such componentry may be integrated or peripheral, and the transduction and/or processing of input actions may be handled on- or off-board. Example NUI componentry may include a microphone for speech and/or voice recognition; an infrared, color, stereoscopic, and/or depth camera for machine vision and/or gesture recognition; a head tracker, eye tracker, accelerometer, and/or gyroscope for motion detection and/or intent recognition.

When included, communication subsystem **510** may be configured to communicatively couple computing system **500** with one or more other computing devices. Communication subsystem **510** may include wired and/or wireless communication devices compatible with one or more different communication protocols. The communication subsystem may be configured for communication via personal-, local- and/or wide-area networks.

In an example, a scale and pose estimation method for a camera system comprises receiving camera data for a scene acquired by one or more cameras of the camera system, receiving a rotation prior parameter characterizing a gravity direction, receiving a scale prior parameter characterizing scale of the camera system, calculating a cost of a cost function for a similarity transformation that is configured to encode a scale and pose of the camera system, such calculation being performed to optimize a rotation term and a translation term of the similarity transformation, where the cost of the cost function is selectively influenced by the rotation prior parameter and the scale prior parameter, determining a solved similarity transformation upon calculating a cost for the cost function that is less than a threshold cost, and outputting an estimated scale and pose of the camera system based on the solved similarity transformation. In this example and/or other examples, the threshold cost may be a minimized cost for the cost function. In this example and/or other examples, the rotation prior parameter and the scale prior parameter may selectively influence the cost of the cost function by selectively imposing a penalty on the cost function. In this example and/or other examples, the rotation prior parameter and the scale prior parameter may be derived from measurements of one or more inertial sensors. In this example and/or other examples, the cost function may include a rotation weight corresponding to the rotation prior parameter and a scale weight corresponding to the scale prior parameter. In this example and/or other examples, the rotation weight and the scale weight may be adjusted based on sensor noise of the one or more inertial sensors. In this example and/or other examples, the rotation weight and the scale weight may decrease as sensor noise increases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have less influence on the cost function as sensor noise increases, and the rotation weight and the scale weight may increase as sensor noise decreases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have more influence on the cost function as sensor noise decreases. In this example and/or other examples, the rotation weight and the scale weight may be set to zero based on sensor noise being greater than a threshold noise level such that the corresponding rotation prior parameter and the corresponding scale prior parameter have no influence on the cost function. In this example and/or other examples, the camera data may include a collection of three-dimensional (3D) points of a 3D model of the scene. In this example and/or other examples, the camera system may include a plurality of different cameras having different positions fixed relative to each other. In this example and/or other examples, the camera system may include a single camera movable throughout the scene.

In an example, a computing system comprises one or more logic machines, one or more storage machines holding instructions executable by the one or more logic machines to receive camera data for a scene acquired by one or more cameras of a camera system, receive a rotation prior parameter characterizing a gravity direction, receive a scale prior parameter characterizing scale of the camera system, calcu-

late a cost of a cost function for a similarity transformation that is configured to encode a scale and pose of the camera system, such calculation being performed to optimize a rotation term and a translation term of the similarity transformation, where the cost of the cost function is selectively influenced by the rotation prior parameter and the scale prior parameter, determine a solved similarity transformation upon calculating a cost for the cost function that is less than a threshold cost, and output an estimated scale and pose of the camera system based on the solved similarity transformation. In this example and/or other examples, the threshold cost may be a minimized cost for the cost function. In this example and/or other examples, the rotation prior parameter and the scale prior parameter selectively influence the cost of the cost function by selectively imposing a penalty on the cost of the cost function. In this example and/or other examples, the rotation prior parameter and the scale prior parameter are derived from measurements of one or more inertial sensors. In this example and/or other examples, the cost function may include a rotation weight corresponding to the rotation prior parameter and a scale weight corresponding to the scale prior parameter. In this example and/or other examples, the rotation weight and the scale weight may be adjusted based on sensor noise of the one or more inertial sensors. In this example and/or other examples, the rotation weight and the scale weight may decrease as sensor noise increases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have less influence on the cost function as sensor noise increases, and the rotation weight and the scale weight may increase as sensor noise decreases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have more influence on the cost function as sensor noise decreases. In this example and/or other examples, the rotation weight and the scale weight may be set to zero based on sensor noise being greater than a threshold noise level such that the corresponding rotation prior parameter and the corresponding scale prior parameter have no influence on the cost function.

In an example, a scale and pose estimation method for a camera system comprises receiving camera data for a scene acquired by one or more cameras of the camera system, receiving a rotation prior parameter characterizing a gravity direction, receiving a scale prior parameter characterizing scale of the camera system, wherein the rotation prior parameter and the scale prior parameter are derived from measurements of one or more inertial sensors, calculating a cost function for a similarity transformation that is configured to encode a scale and pose of the camera system, such calculation being performed to optimize a rotation term and a translation term of the similarity transformation, where the cost of the cost function is selectively influenced by the rotation prior parameter, a rotation weight corresponding to the rotation prior parameter, the scale prior parameter and a scale weight corresponding to the scale prior parameter, adjusting the rotation weight and the scale weight based on the sensor noise of the one or more inertial sensors, determining a solved similarity transformation upon calculating a cost for the cost function that is less than a threshold cost, and outputting an estimated scale and pose of the camera system based on the solved similarity transformation.

This disclosure is presented by way of example and with reference to the associated drawing figures. Components, process steps, and other elements that may be substantially the same in one or more of the figures are identified coordinately and are described with minimal repetition. It will be noted, however, that elements identified coordinately

may also differ to some degree. It will be further noted that some figures may be schematic and not drawn to scale. The various drawing scales, aspect ratios, and numbers of components shown in the figures may be purposely distorted to make certain features or relationships easier to see.

It will be understood that the configurations and/or approaches described herein are exemplary in nature, and that these specific embodiments or examples are not to be considered in a limiting sense, because numerous variations are possible. The specific routines or methods described herein may represent one or more of any number of processing strategies. As such, various acts illustrated and/or described may be performed in the sequence illustrated and/or described, in other sequences, in parallel, or omitted. Likewise, the order of the above-described processes may be changed.

The invention claimed is:

1. A scale and pose estimation method for a camera system, the method comprising:

receiving camera data for a scene acquired by one or more cameras of the camera system;  
receiving a scale prior parameter characterizing scale of the camera system;  
calculating a cost of a cost function for a similarity transformation that is configured to encode a scale and pose of the camera system, where the cost of the cost function is selectively influenced at least by the scale prior parameter; and  
based at least on the cost of the cost function being less than a threshold cost, outputting an estimated scale and pose of the camera system based on the similarity transformation.

2. The method of claim 1, wherein the threshold cost is a minimized cost for the cost function.

3. The method of claim 1, wherein the calculation of the cost of the cost function is performed to optimize a rotation term and a translation term of the similarity transformation.

4. The method of claim 1, further comprising receiving a rotation prior parameter characterizing a gravity direction, wherein the cost of the cost function is selectively influenced at least by the rotation prior parameter.

5. The method of claim 4, wherein the rotation prior parameter and the scale prior parameter selectively influence the cost of the cost function by selectively imposing a penalty on the cost function.

6. The method of claim 4, wherein the rotation prior parameter and the scale prior parameter are derived from measurements of one or more inertial sensors.

7. The method of claim 6, wherein the cost function includes a rotation weight corresponding to the rotation prior parameter and a scale weight corresponding to the scale prior parameter.

8. The method of claim 7, wherein the rotation weight and the scale weight are adjusted based on sensor noise of the one or more inertial sensors.

9. The method of claim 8, wherein the rotation weight and the scale weight decrease as sensor noise increases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have less influence on the cost function as sensor noise increases, and wherein the rotation weight and the scale weight increase as sensor noise decreases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have more influence on the cost function as sensor noise decreases.

10. The method of claim 8, wherein the rotation weight and the scale weight are set to zero based on sensor noise being greater than a threshold noise level such that the

corresponding rotation prior parameter and the corresponding scale prior parameter have no influence on the cost function.

11. The method of claim 1, wherein the camera data includes a collection of three-dimensional (3D) points of a 3D model of the scene.

12. A computing system comprising:  
 one or more logic machines;  
 one or more storage machines holding instructions executable by the one or more logic machines to:  
 receive camera data for a scene acquired by one or more cameras of a camera system;  
 receive a scale prior parameter characterizing scale of the camera system;  
 calculate a cost of a cost function for a similarity transformation that is configured to encode a scale and pose of the camera system, where the cost of the cost function is selectively influenced at least by the scale prior parameter; and  
 based at least on a cost of the cost function being less than a threshold cost, output an estimated scale and pose of the camera system based on the similarity transformation.

13. The computing system of claim 12, wherein the threshold cost is a minimized cost for the cost function.

14. The computing system of claim 12, wherein the instructions are executable by the one or more logic machines to receive a rotation prior parameter characterizing a gravity direction, wherein the cost of the cost function is selectively influenced at least by the rotation prior parameter.

15. The computing system of claim 14, wherein the rotation prior parameter and the scale prior parameter selectively influence the cost of the cost function by selectively imposing a penalty on the cost of the cost function.

16. The computing system of claim 14, wherein the rotation prior parameter and the scale prior parameter are derived from measurements of one or more inertial sensors.

17. The computing system of claim 15, wherein the cost function includes a rotation weight corresponding to the rotation prior parameter and a scale weight corresponding to the scale prior parameter, and wherein the rotation weight

and the scale weight are adjusted based on sensor noise of the one or more inertial sensors.

18. The computing system of claim 17, wherein the rotation weight and the scale weight decrease as sensor noise increases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have less influence on the cost function as sensor noise increases, and wherein the rotation weight and the scale weight increase as sensor noise decreases such that the corresponding rotation prior parameter and the corresponding scale prior parameter have more influence on the cost function as sensor noise decreases.

19. The computing system of claim 17, wherein the rotation weight and the scale weight are set to zero based on sensor noise being greater than a threshold noise level such that the corresponding rotation prior parameter and the corresponding scale prior parameter have no influence on the cost function.

20. A scale and pose estimation method for a camera system, the method comprising:  
 receiving camera data for a scene acquired by one or more cameras of the camera system;  
 receiving a rotation prior parameter characterizing a gravity direction;  
 receiving a scale prior parameter characterizing scale of the camera system, wherein the rotation prior parameter and the scale prior parameter are derived from measurements of one or more inertial sensors;  
 calculating a cost function for a similarity transformation that is configured to encode a scale and pose of the camera system, where the cost of the cost function is selectively influenced by the rotation prior parameter, a rotation weight corresponding to the rotation prior parameter, the scale prior parameter and a scale weight corresponding to the scale prior parameter;  
 adjusting the rotation weight and the scale weight based on the sensor noise of the one or more inertial sensors;  
 based at least on a cost of the cost function being less than a threshold cost, outputting an estimated scale and pose of the camera system based on the similarity transformation.

\* \* \* \* \*